

History and Context of the NAEP Achievement Levels

One of the Governing Board's most important legislative responsibilities is developing the NAEP achievement levels. The history of NAEP achievement level setting (also known as standard setting) over the past 30 years is complex.

During the 1980s, one aspect of NAEP reporting was the use of anchor or performance levels where NAEP scales were divided into five regions that reflected performance at the mean and one and two standard deviations below and above the mean. This type of norm-referenced approach to performance levels is still used by some of the international assessments, but the NAEP legislation specifies that the NAEP achievement levels be based on a "national consensus approach," which clearly indicates the role of human judgment to compare performance to explicit goals rather than simply reflecting how students perform relative to others.

When the Governing Board developed its initial policy statement for developing NAEP achievement levels in 1990, the field of standard setting was much less developed than it is today and most experiences were in the area of professional licensure. At that time, very little standard setting had been conducted in K-12 education, particularly at the national level. The Governing Board's initial efforts to develop achievement levels for NAEP were thus groundbreaking, and the Board's research and practices in this area over the past three decades have had a large impact on how standard setting is conducted today in K-12 education in the United States.

Early evaluations of the NAEP achievement levels (conducted in the 1990s) were critical of the Governing Board's procedures and of judgmental standard setting in general. When the initial NAEP legislation was re-authorized in 1994, it included language that the NAEP achievement levels should be "used on a developmental basis until the [NCES] Commissioner determines, as a result of an evaluation... that such levels are reasonable, valid, and informative to the public" (Public Law 103-382). The next re-authorizations substituted the term "trial" for "developmental" (Public Laws 107-110 and 107-279).

In 1999, a [report of an evaluation conducted by the National Academy of Sciences](#) stated that the process for setting NAEP achievement levels was "fundamentally flawed." This conclusion was disputed by many technical experts, and several prominent researchers issued a [response](#) criticizing the evaluation. Although the Governing Board has adopted new approaches to developing the NAEP achievement levels since the 1990s, the original methodology (called "modified Angoff") is still considered a reputable approach to standard setting and remains in wide use today. For example, a version of the modified Angoff methodology was used to [set the achievement levels for the Partnership for Assessment of Readiness for College and Careers \(PARCC\)](#).

The more recent evaluations of the NAEP achievement levels have recognized their value and have been more positive about the standard setting methodology used by NAEP. The [2009 evaluation of the NAEP program](#) as a whole included a focus on the Mapmark achievement-level setting activities for the 2005 grade 12 mathematics assessment and noted:

“Based on our analysis of the procedural, internal, and external evidence pertaining to the validity of the process of setting achievement level standards on the 2005 Grade 12 NAEP Mathematics assessment, we conclude that the procedure was sound, followed recommendations for best practices in the area of standard setting, and involved multiple quality control checks to support the defensibility of the process. The validity of any type of test score interpretation is not something that can be unequivocally established. However, the multiple sources of validity evidence we analyzed for this exam lead us to conclude the standards set on the 2005 Grade 12 NAEP Mathematics assessment are valid for the purposes of describing the performance of 2005 Grade 12 students with respect to the NAEP achievement level descriptors” (pages 2-58 to 2-59).

The report included several recommendations to further improve the process.

The [2016 evaluation of NAEP achievement levels](#), conducted by the National Academies of Sciences, Engineering, and Medicine, focused on the NAEP mathematics and reading achievement levels for grades 4, 8, and 12. This evaluation report, presented to the Governing Board at its November 2016 meeting, stated, “During their 24 years [the achievement levels] have acquired meaning for NAEP’s various audiences and stakeholders; they serve as stable benchmarks for monitoring achievement trends, and they are widely used to inform public discourse and policy decisions. Users regard them as a regular, permanent feature of the NAEP reports” (page Sum-8). This evaluation included several recommendations, and the Board issued a formal response noting its planned actions in December 2016.

The evaluation recommended that research be undertaken to review and potentially revise the achievement level *descriptions* for reporting (an update to studies the Board previously conducted in 2009¹). The evaluation did not conclude that the NAEP achievement *levels* should be revisited, reset, or abandoned. The Governing Board has recently issued a contract to Pearson for performing the work to review and revise the achievement level descriptions. In addition, the evaluation included several recommendations for ways that the Board can work (in collaboration with NCES) to improve the interpretation and communication of achievement levels.

¹ The term “achievement levels” refers to categories of performance that are created when cut scores are applied to an assessment (i.e., NAEP Basic, NAEP Proficient, and NAEP Advanced). The Board policy includes definitions of the achievement levels that apply to all subjects and grades). In addition, there are “achievement level descriptions” which translate the policy definitions into specific expectations about student knowledge and skills in a particular content area, for each subject and grade (e.g., grade 4 mathematics).

One important aspect of the Board's response to the 2016 evaluation was a commitment to update the guidance provided in the Board policy statement on NAEP achievement level setting. The Committee on Standards, Design and Methodology (COSDAM) worked to update this policy guidance over a period of 1.5 years, which included convening expert panels, commissioning white papers, and performing literature reviews on current best practices in standard setting, including what could be learned from state assessments and other programs. These research efforts and their implications for revising the policy were discussed at nearly every COSDAM meeting during 2017 and 2018. Finally, the Board sought and considered public comment on a draft policy statement and unanimously adopted a revised Board policy on [Developing Student Achievement Levels for NAEP](#) (attached) in November 2018.

Following the adoption of the revised policy, the Board formed an Achievement Levels Working Group at the March 2019 Board meeting, led by Board member and standard setting expert Gregory Cizek. The purpose of the Achievement Levels Work Plan was to describe how the Board's initial response to the evaluation will be implemented, using the newly revised policy, in terms of specific activities and expected timelines². It is intended to be a collaborative effort between the Board and NCES, particularly related to the efforts for improving the reporting and communication of the NAEP achievement levels. The [Achievement Levels Work Plan](#) (attached) was adopted at the March 2020 Board meeting and amended during the July 2020 Board meeting. Under the leadership of COSDAM, the Board will continue to pursue the activities outlined in this plan.

² The COVID-19 pandemic has impacted the proposed timeline for some of the planned activities that will be pursued by both the Governing Board and NCES, due to unanticipated challenges and the inability to conduct in-person activities.